CHAPTER 4
TRANSFORM DOMAIN BASED VISUAL SALIENCY DETECTION

4.1 INTRODUCTION

This chapter focuses on the visual saliency detection method based on Ripplet transform. It explores the detection of visual saliency in transform domain. The computational models for visual attention are derived in the spatial domain as well as in the frequency domain. The spatial domain models suffer from high computational complexity (Ru-Je Lin & Wei-Song Lin 2014) and are not suited for multi scale analysis. The frequency domain methods are more stout against noise and effectively represent images compared to the spatial domain. In the recent past, the researchers were very much interested to build the computational models in the frequency domain.

The proposed methodology for saliency detection uses Ripplet transform which has unique advantages compared to the other transforms. The transform domain methods generally undergo the following steps: Obtain the local and global features. Apply the corresponding transforms. Find the various feature maps. Combine the feature maps in order to get the final saliency map.

The organization of the chapter is as follows: Section 4.2 highlights the different transforms which are used in the visual saliency detection. Section 4.3 describes the Ripplet transform. Section 4.4 explains the proposed
methodology for saliency detection using Ripplet transforms. Section 4.5 shows the experimental results of the proposed method and state-of-the-art methods under various datasets. It also presents the performance metrics of the proposed method and state-of-the-art methods. Summary of the chapter is given in the section 4.6

4.2 TRANSFORMS IN THE VISUAL SALIENCY DETECTION

The major transforms which are used in the visual saliency detection methods are Fourier transform, phase spectrum of quaternion Fourier transform, discrete cosine transforms, wavelet transform, Curvelet transform and Shearlet transform. Most of the literatures in the field of saliency detection used Fourier transform and wavelet transforms.

The phase spectrum, magnitude spectrum, hyper-complex magnitude spectrum, hyper-complex phase spectrum of Fourier transform is used to develop the computational models for visual attention. Apart from this the Fourier transform is used along with the log spectrum and Quaternion Discrete Cosine Transform. Phase Spectrum of Quaternion Fourier Transform (PQFT) is also used to obtain the visual saliency detection methods. Fourier transform has some limitations. It could not give satisfactory results for the aperiodic signals and non-stationary signals. It has the problem to represent the local frequency components. In order to address these problems, the researchers shift to wavelet transform.

In comparison with the Fourier transform, the wavelet transform has very good capabilities. It can represent the signal at various bands and bandwidths. So the multi scale spatial and frequency analysis is done through wavelet transforms. The point singularities are well identified by the wavelet transforms. So the wavelet coefficients along with the entropy are used to
develop the computational models of visual attention. However, the wavelet transforms also has some limitations.

Wavelet transform fails to represent the intermediate dimensional structures. Even though the point singularities are represented well by wavelet transforms, it fails to represent the directional features. It is due to the isotropic elements of wavelet transform. While approximating the curve using wavelet transform, the level of decomposition is very high. So the directional features are not represented well by wavelet transform. The multi-directional and multi-scale transforms are utilized to overcome these disadvantages. The quantity of decomposition level should be very large when approximating a curve using WT. The disadvantages of WT are overcome by using multi-directional and multi-scale transforms.

The two dimensional singularities of line and curve are well expressed by the Curvelet transform. It is nothing but multi-scale Ridgelet transform. The visual saliency detection method is proposed with the help of Curvelet transform and 2D Gabor transform (Sheng-hua Zhong 2011). In order to enhance the geometric features, Shearlet and platelet transforms are developed. Shearlet transform based visual saliency detection methods were also proposed in the recent past literatures (Lei Bao et al. 2014). The scaling law is followed by the Curvelet transform in which the anisotropic property is the key parameter to represent the 2D singularities. Width (W) = Length^2 (L^2) is the parabolic scaling law. It is not ideal for the all types of boundaries. To satisfy this constraint, the scaling law is refined

The general form of Curvelet transform sharpen scaling law is called Ripplet transform. Support c and degree d are the two parameters which are used in the Ripplet transform. Anisotropy capabilities of Ripplet transform are used to represent the higher dimensional singularities. The existing methods of saliency detection which are used in the higher
dimensional wavelets fail to represent the directional features in an effective way.

The directional features of an image are well represented by Ripplet transform. It is also has some unique properties such as general scaling and support, anisotropy, multi resolution, good localization, high directionality and fast coefficient decay. These properties are essential to describe the images for saliency detection. So in this proposed methodology Ripplet transform is used for visual saliency detection.

4.3 RIPPLET TRANSFORM

Xu et al. (2010) proposed the Ripplet Transform (RT) which breaks the ingrained difficulties of Wavelet transform to the utmost level. It has the capability to represent images at different scales and different directions. This higher dimensional Wavelet transform is optimal to represent $\mathcal{C}^2$ singularities. In Ripplet space, edges of an image have a sparse representation. The Ripplet coefficients of an image are generated by convolving the Ripplet with an image.

The Ripplet function can be defined as:

$$\rho_{ab\theta}(\vec{x}) = \rho_{a\vec{0}}(R_{\theta}(\vec{x} - \vec{b}))$$  \hspace{1cm} (4.1)

where $R_{\theta} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$ is the rotation matrix which rotates $\theta$ radians. The set of functions $\{\rho_{ab\theta}\}$ is defined as Ripplet functions or Ripplets because in spatial domain these functions have ripple-like shapes (Xu et al. 2010). $\vec{x}$ and $\vec{b}$ are 2D vectors. $\rho_{a\vec{0}}(.)$ is the mother function of Ripplet in frequency domain (Sujitha Juliet et al. 2013).
Digital image processing uses discrete transforms instead of continuous transform. The discretization process (Xu et al. 2010) of continuous Ripplet transform is based on the discretization of Ripplet parameter. The frequency response of Ripplet function in frequency domain is given by:

\[ \hat{\rho}_j(r, \omega) = \frac{1}{\sqrt{c}} a \frac{m+n}{2n} W(2^{-j}.r)V \left( \frac{1}{c} . 2^{-\left\lfloor \frac{m-n}{n} \right\rfloor} \omega - l \right) \]  \hspace{1cm} (4.2)

where \( W \) and \( V \) are the ‘radial window’ and ‘angular window’ respectively and should satisfy the following admissibility conditions:

\[ \sum_{j=0}^{+\infty} |W(2^{-j}.r)|^2 = 1 \]  \hspace{1cm} (4.3)

\[ \sum_{l=-\infty}^{+\infty} |V(\frac{1}{c} . 2^{-\left\lfloor \frac{1}{d} \right\rfloor} \omega - l)|^2 = 1 \]  \hspace{1cm} (4.4)

The scale parameter \( (a) \) is sampled at dyadic intervals. The position parameter \( (b) \) and rotation parameter \( (\theta) \) are sampled (Xu et al. 2010) at equally-spaced intervals.

\[ a_j = 2^{-j}, \overline{b}_k = [c \times 2^{-j} \times k_1, 2^{-j} \times k_2]^T \text{ and } \theta_l \]

\[ = \frac{2\pi}{c} \times 2^{-\left\lfloor \frac{1}{d} \right\rfloor} \times l \]  \hspace{1cm} (4.5)

where \( \overline{k} = [k_1, k_2]^T \) and \( j, k_1, k_2, l \in Z \times (.)^T \) denote transpose of a vector, \( d \in R \), it can be approximated by rational numbers represented by rational numbers as \( d = n/m \), for all \( n, m \neq 0 \in Z \) and \( n, m \) are primes. The Discrete Ripplet transform (DRT) of \( M \times N \) image \( f(n_1, n_2) \) will be in the form as shown below:
\[ R_{j,k,l} = \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{N-1} f(n_1, n_2) \rho_{j,k,l}(n_1, n_2) \]  

where \( R_{j,k,l} \) are the Ripplet coefficients. The image can be reconstructed through Inverse Discrete Ripplet transform

\[ \tilde{f}(n_1, n_2) = \sum_j \sum_k \sum_l R_{j,k,l} \rho_{j,k,l}(n_1, n_2) \]  

The properties of Ripplet transform are used in various applications like image steganography, image fusion, image compression and image retrieval. Uma Maheswari & Jude Hemanth (2015) proposed Ripplet transform based image steganography. It is used to represent the images in multi scale and multi direction in order to embed the secret data.

Multi scale geometric analysis of Ripplet transform is used in the natural color image retrieval system (Manish Chowdhury et al. 2013). The directionality property of Ripplet transform is utilized by Peng Geng et al. (2014) for image fusion applications. It is also used in the medical image compression by Sujitha Juliet et al. (2013). High quality compressed medical images are obtained by Ripplet transform.

**4.4 PROPOSED METHODOLOGY FOR RIPPLET TRANSFORM BASED VISUAL SALIENCY DETECTION**

All the types of images comprise important and unimportant regions. Visual saliency is noted as unusual regions from the background of an image. The unimportant regions are developed by the image feature such as color, texture, orientation and shape. The computational models of visual attention highlight the salient regions by considering various conditions. The two dimensional singularities of edges and textures are well represented by
Ripplet transform. The ripplet coefficients are effectively representing the uncommon regions (Xu et al. 2010). The uncommon regions are different compared to the image background. These salient regions are fetched out by using Ripplet transform which represents image features in different directions and different scale. The variation between unusual regions and their background area is effectively highlighted by the Ripplet transform. It has also the capability to represent orientation singularities (Sujitha Juliet et al. 2013). So the proposed methodology is used with Ripplet transform for visual saliency detection.

![Image of original and transformed ripplet patterns](image)

**Figure 4.1** Original image patterns (top row) vs. Transformed image patterns in Ripplet domain (bottom row)

Figure 4.1 shows the various image patterns and their decomposed representations in the Ripplet domain. First row shows the original image patterns. The uncommon regions of original images are highlighted by their basic image features such as, orientation, density, color, texture and shape which are shown in the second row. From the Figure 4.1, it is understood that the Ripplet transform is connected with the detection of saliency with the proper image features. The salient regions can be identified through Ripplet transform if the color channels, orientation, scale are chosen correctly.

The frame work of the proposed methodology is shown in the Figure 4.2. An efficient visual saliency detection model based on Ripplet
transform is proposed. The input images are represented in different scale and different directions by the Ripplet coefficients. The feature maps are obtained by multi-level decomposition with inverse Ripplet transform. The local and the global saliency maps are obtained with the help of image contrast and entropy. At last the final saliency map is furnished by combining both the saliency maps.

4.4.1 Color Channels

Most of the visual saliency detection methods rely on the contrast of the input images. The image contrast is one of the local features. The images are represented with different contrast in different color channels. So the proper color channels need to be selected in the saliency analysis (Luo et al.)
The existing visual saliency detection methods used different color channels such as Lab, YCbCr and RGBYI. In most recent the RGBYI color channel (Borji 2012) is introduced in existing visual saliency detection literatures.

4.4.2 Feature Map Generation

Let \( I^{cl} \) be the color image. In the preprocessing stage input is applied with two dimensional \( n \times n \) Gaussian low pass filter to remove the noise.

\[
I = I^{cl} * G_{n \times n} \tag{4.8}
\]

where \( G \) is the two dimensional \( n \times n \) Gaussian low pass filter \((n = 3)\). Consider \( I \) is the noise free image which is obtained by performing convolution with input color image \( I^{cl} \) and two dimensional \( n \times n \) Gaussian low pass filter. Each color channel is done with the normalization of \([0,255]\). The ganglion cells of human retina system are very much sensitive to color, texture and orientation (Kaplan et al. 1986). It is mocked by the Ripplet transform (Manish Chowdhury et al. 2013).

Discrete Ripplet transform is applied to input image to get image sub-bands. Four level (N=4) Ripplet decomposition is done \((1,2,4,4)\). Totally 11 sub-bands for each color channel is obtained from ripplet decomposition. 99 sub-bands are obtained from the nine color channels of an image.

\[
R_{pq}^c = DRT_N(I) \tag{4.9}
\]

where \( c \) is the channel of the noise free image \( I \).

\[
c \in \{r,g,b,I,RED,GREEN,BLUE,RGE,BYL\};
\]
The details of the noise free image are represented through ripplet coefficients $R_{pq}^c$ is the ripplet coefficients at $p$ scale and $q$ direction. (k=8 is taken for each level). The images are represented in multi-scale and multi-direction by the ripplet coefficients. By applying Inverse Ripplet transform (IRT) the feature maps $f_k^c(x, y)$ are obtained. It is obtained for k level decomposition for individual sub-bands. i.e $k \in (1....N)$. There are large numbers of feature values are existing. In order to limit the feature values the scaling factor $\beta$ is introduced as $10^3$

$$f_k^c(x, y) = \{IRT[R_{pq}^c]\}^2/\beta \quad (4.10)$$

### 4.4.3 Global Saliency Map

The computation of saliency map is the very next step of feature maps. Two types of saliency maps are obtained. First one is global saliency map in which the local features and global distributions are considered. At the location (x,y) the feature vector is $f(x,y)$. It is having the length of $9 \times N$ and it is described in the Equation (4.8). At each location (x,y), totally 36 features are taken. Imamoglu et al. (2013) method is adopted to obtain the global saliency map. Probability Density Function (PDF) and normal distribution are used to define features likelihood which are in the feature maps.

Feature vector

$$f(x, y) = [f^c(x, y), \ldots, f_N^c(x, y)]^T \quad (4.11)$$

The Gaussian PDF in multi-dimension is given by

$$p(f(x, y)) = \frac{1}{2\pi n/2|\Sigma|^{1/2}} \times e^{-1/2(f(x,y)-\mu)^T \Sigma^{-1}(f(x,y)-\mu)} \quad (4.12)$$
where $\Sigma = E[(f(x, y) - \mu)(f(x, y) - \mu)^T]$ is $n \times n$ covariance matrix ($n = 9 \times N$).

$|\Sigma|$ is the determinant of the covariance matrix.

$\mu$ is the mean vector of each feature map; $\mu = E[f]$.

$T$ is the transpose operation.

Equation (4.9) is the PDF of feature maps. The global statistics of feature maps mean and covariance matrix are considered as global outcome. Two dimensional gaussian low pass filter $H_{k \times k}$ is used to achieve the results smoothly ($k=5$).

$$S_{gbat}(x, y) = (\log(p(f(x, y))^{-1}))^{0.5} * H_{k \times k} \quad (4.13)$$

where $S_{gbat}(x, y)$ is the global saliency map.

The global saliency map highlights the statistical relationship between feature maps. It also showcase the vital information which is not showcased by the local contrast. Equation (4.13) yields the saliency map in which some of the local information are missed out. It is due to the structure of the input image.

Global saliency maps for the color images are displayed in the Figure 4.3. It shows how the salient regions are highlighted by the local features. Local contrast is balanced with the local features distribution. Figure 4.3(a) shows the input color images Figure 4.3(b) shows the global saliency map and Figure 4.3(c), (d) shows the local saliency map and final saliency map. The images of Figure 4.3 are having smooth background. So the salient regions are curbed by the saliency maps.
Figure 4.3 Smooth background Images and their local and global saliency maps (a) color images (b) global saliency map (c) local saliency map (d) final saliency map

Majority of the natural images are not having smooth backgrounds. So it is needed to consider the non-smooth background images also. In some images the local contrast is fully suppressed by the global saliency map (refer top row of Figure 4.4(b) and Figure 4.4(c)). In some images local saliency identifies potential salient regions but which are not considered or less considered globally. In another set of images the global saliency manifests the potential salient regions but they are not considered or less considered locally. It is shown in the Figure 4.4.

The bird image is shown in the Figure 4.4 in which the head portion and wing portion are having high local saliency (Figure 4.4(c)). But in Figure 4.4(b) the local features and global distributions are highly considered. The wings are given higher prominence than the head of the bird (refer bottom row Figure 4.4(b))

Therefore it is understood that the efficient saliency maps can be obtained by utilizing different saliency maps of local features. So it is necessary to consider the local features global information.
4.4.4 Local Saliency Map

Entropy plays vital role to find the amount of information in the scene. In this proposed method, the concept of entropy is used to get the local saliency map. Usually image local areas consist of salient regions (Ru Je Lin et al. 2014). While comparing the well-organized regions with disorganized regions of an image, the entropy of well-organized regions is lower than that of the disorganized regions (Duncan et al. 2012). The disorganized regions are having higher entropy and lesser saliency value. If the local areas of an image are very much closer to each other, then the entropies of the local area become larger. It seems none of the local areas are salient. But in some cases some of the local areas are extremely picked out from their backgrounds, their entropies are lower and their saliencies are higher. This idea is leveraged in the local saliency computation. For a test case, assume there is a beautiful white dove rambling around a green grass field. In this moment, the beautiful white dove is interpreted as saliency. The global and local contrasts are used to detect the corresponding saliency maps. The contrasts are types of texture information only. Suppose the green grass has rough surface and the white dove has very smooth feathers, then the green grass has much chance of being...
viewed as salient. So the contrast and color differences of local areas play an important role in the visual saliency detection.

Entropy is reflecting the energy of an input image but it could not give the energy distribution. The proposed method uses entropy and feature distribution for the effective local saliency map generation. Figure 4.3(d) and Figure 4.4(d) show the examples of proposed saliency map. It is noted that there are some differences between local saliency map (Figure 4.3(c) and Figure 4.4(c)) and global saliency map (Figure 4.3(b) and Figure 4.4(b)).

The feature distribution is updated by probability of saliency of a given feature map $f(x, y)$ in a given color image $I^c$. The feature distribution is done by considering the information theory and pixelwise similarity. Similarity function is introduced to enhance the probability of features which are appearing in the image. $Y_s^l$ is the similarity function of $l^{th}$ feature in ’s’ position given by Equation (4.14)

$$Y_s^l = \frac{1}{W \times Y} \sum_{i \in I^c, i \neq s} Q(b|f_m^l - f_n^l|+1)$$  \hspace{1cm} (4.14)

$$Q(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}$$  \hspace{1cm} (4.15)

The size of the input image is represented by $W \times Y$ and b is constant 

(b = -0.01). The smaller contrast between $f_m^l$ and $f_n^l$ yields higher value for $Q$. $Y_s^l$ is observed as appearance probability (Ru Je Lin et al. 2014) of $l^{th}$ feature in $s$ location. It is improved by all other comparable features in that location of the image $I^c$. 
The appearance probability is determined for multi-level because the saliency of the proposed method is quantified for multi-scale. Entropy $\varepsilon_N(x, y)$ is obtained for the local areas with the neighborhood of $(k + 1) \times (k + 1)$. $k$ is considered as 5 in this work. The saliency at the location of $(x, y)$ is found by,

$$\varepsilon_S^c(x, y) = -\sum_{i=1}^{2k+1} Y_S^l(x_i y_i) \times \log_{10} Y_S^l(x_i y_i) \quad (4.16)$$

$$\varepsilon_S(x, y) = \sum_c N( \varepsilon_S^c (x, y)) \quad (4.17)$$

$N(.)$ is the normalization operator. $\varepsilon_S^c(x, y)$ is incorporated for every location of $(x, y)$ to find the local saliency map $S_{local}$.

$$S_{local} = (\sum_{s=1}^{N} \varepsilon_s (x, y)) \ast g_{n \times n} \quad (4.18)$$

$S_{local}$ is the local saliency map.

### 4.4.5 Final Saliency Map

The final saliency map is obtained by combining the global saliency map (Equation 4.13) and local saliency map (Equation 4.18) (Imamoglu et al. 2013)

$$S_{fl}(x, y) = M(S_{local}(x, y) \times e^{S_{global}(x, y)}) \quad (4.19)$$

$M(\cdot) = (\cdot)^{ln/2}/\sqrt{2}$. $M$ is the modulation which is used to decay the amplification effect on the final saliency map. The saliency values which are nearer the salient points, are enhancing the performance of the saliency map (Goferman et al. 2010). The focused attention points have more impact than the points away from the attention. The values which are greater than 0.7 are considered as focus of attention points in this proposed method.
M(\cdot) = (\cdot)^{\ln\sqrt{2}/\sqrt{2}}. Goferman et al. (2010) states that the saliency values around the salient points are boosting the enhancement of the performance of saliency map. The focus of attention has more impact than those far away from the attention. In this proposed method, saliency values greater than 0.7 are assigned as focus of attention points.

\[ s(x, y) = S_f(x', y')(1 - d_{FOA}(x, y)) \]  \hspace{1cm} (4.20)

\( s(x, y) \) is the saliency value at the location of \((x, y)\). \((x', y')\) is the most salient location. The saliency value at the location \((x', y')\) is given by \(S_f(x', y')\). The distance between location \((x, y)\) and its nearest FOA at the location \((x', y')\) is represented by \(d_{FOA}(x, y)\). In the focused region the saliency values are high, which will be reflected in the final saliency map also.

4.5 EXPERIMENTAL RESULTS

The experimental results section explains the results of the proposed method and state-of-the-art methods.

4.5.1 Experimental Setup

The experiment is conducted for various databases which are used in the literatures of visual saliency detection.

MSRA-5000: MSRA-5000 is having 5,000 natural images and the ground truths of (Liu et al. 2011) corresponding images. All the images are having simple background with single salient object in the foreground. The ground truths are specified with human labeled rectangles around the salient objects. Different types of images are presented in this dataset. Indoor, outdoor, natural scenes, animals etc. are presented in this dataset.
The specific characteristics of each benchmark database are shown in the Table 4.1.

### Table 4.1 List of benchmark datasets used

<table>
<thead>
<tr>
<th>Name</th>
<th>Size (no of images)</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSRA 5000</strong> (Liu et al. 2011)</td>
<td>5000</td>
<td>Single object, simple background, high contrast</td>
</tr>
<tr>
<td><strong>MSRA 1000</strong> (subset of MSRA 5000) (Achanta et al. 2009)</td>
<td>1000</td>
<td>Single object, simple background, high contrast</td>
</tr>
<tr>
<td><strong>SED1</strong> (Alpert S et al. 2007)</td>
<td>100</td>
<td>Single foreground object</td>
</tr>
<tr>
<td><strong>SED2</strong> (Alpert et al. 2007)</td>
<td>100</td>
<td>Two foreground objects</td>
</tr>
<tr>
<td><strong>SOD</strong> (Movahedi et al. 2010)</td>
<td>300</td>
<td>Multiple objects, complex background</td>
</tr>
</tbody>
</table>

#### 4.5.2 Performance Analysis

The experimental work of the proposed method is carried out for the database shown in the Table 4.1. The results of the proposed methodology are compared with ten state-of-the-art method. The most relevant methods are only taken for performance comparison. The comparison methods are namely IT (Itti et al. 1998), VS (Li et al. 2013), SR (Hou et al. 2007), SE (Murray et al. 2011), WT (Imamoglu et al. 2013), ST (Lei Bao et al. 2014), WE (Xiaolong Ma et al. 2015), PQFT (Chenlei Guo et al. 2010), HSC (Ce Li et al. 2013) and DQCT (Boris et al. 2012). The reasons for choosing the comparison methods are explained below.

IT is the very first method in the scenario of visual saliency detection. Rests of the nine methods are experimented in different transform domains. The methods of PQFT (Chenlei Guo et al. 2010), HSC (Ce Li et al.
VS (Ce Li et al. 2013) and SR (Hou et al. 2007) are Fourier transform based techniques. The methods of SE (Murray et al. 2011), WT (Imamoglu et al. 2013) and WE (Xiaolong Ma et al. 2015) are Wavelet transform based techniques. The method of DQCT (Boris et al. 2012) is the cosine transform and Fourier transform based technique. The method of ST (Lei Bao et al. 2014) is the Shearlet transform based technique. In order to perform viable comparison, the saliency maps are obtained by the ten state-of-the-art methods and normalized in the range of [0,255]

The validation of proposed method is performed with the help of performance metrics. Receiver Operating Characteristic (ROC) curve is taken as one of the performance metrics. The saliency map always consists of salient regions and non-salient regions. ROC is the curve which is drawn between True Positive Rate (TPR) and False Positive Rate (FPR). True positive rate is the amount of target points in the ground truth falling into the salient points of the saliency map. The percentage of the background points falling into the salient points is called False Positive Rate (FPR) (Imamoglu et al. 2013). Each point of the ROC represents trade-off between false positives and false negatives. ROC curve is plotted in the Figure 4.10 which demonstrates the efficacy of the proposed RT method for improving the saliency detection performance.

Further analysis is performed under ROC curve. It is represented by Area Under Curve (AUC). If the AUC values closely approach one then, it is perfect prediction. Figure 4.9 shows the AUC of the proposed and state-of-the-art methods. It clearly exhibits the AUC output of the proposed method in comparison with state-of-the-art methods. The proposed method effectively contributes towards saliency detection with improved performance under all the five databases.
Figure 4.5 Examples of saliency maps over MSRA-1000 dataset
(a) Input images (l) proposed RT method (m) Ground Truths
and (b) – (k) saliency maps of other existing methods

Figure 4.6 Examples of saliency maps over MSRA 5000 dataset
(a) input images (l) proposed RT method (m) Ground Truths (GT) and (b)–(k) saliency maps of other existing methods
Figure 4.7  Examples of saliency maps over SED 1(row 1-3) and SED 2(row 4-6) datasets (a) input images (l) proposed RT method (m) Ground Truths and (b)–(k) saliency maps of other existing methods

Figure 4.8  Examples of saliency maps over SOD dataset: (a) input images, (b)–(k) saliency maps of other existing methods, (l) proposed RT method and (m) ground truth
4.5.2.1 Subjective evaluation

The subjective evaluations of the proposed method RT and state-of-the-art methods are performed on five benchmark databases. The Figures 4.5-4.8 are portrayed with the saliency maps of the proposed method RT and state-of-the-art methods. It is observed that if the images have simple background and homogenous objects, then high quality saliency maps are obtained by different state-of-the-art-methods (such as row 3 and row 4 of Figure 4.5). The proposed RT method highlights the salient region even though the image backgrounds are cluttered (see the Figure 4.7 row 2 and 4, Figure 8 row 1 and 3, Figure 4.6 row 4 and 5). The images with heterogeneous objects are also handled well by the proposed RT method (example automobiles in Figure 4.7 and Figure 4.8, and a person in Figure 4.6 and Figure 4.8). It is inferred that the images with complex background can be effectively handled by the proposed RT method.

Images with multiple objects are also effectively detected by the proposed RT method (see Figure 4.8, row 1,2 and 4, Figure 4.7 last row). The state-of-the-art methods of SE, ST and PQFT do not handle multiple object images effectively. So the proposed RT method proves its applicability in such scenarios also. Large scale (see Figure 4.7 row 2, Figure 5 last row, Figure 4.6 row 1 and 5) and small scale salient object (see Figure 4.7 last row) images are also handled effectively by the proposed RT method. The subjective evaluation of the proposed RT method with other comparison methods clearly highlights that the proposed method successfully handles the wide range of images such as heterogeneous objects, homogenous objects, multiple objects, simple background and cluttered background.
Figure 4.9 AUC of the proposed RT method and other existing method

Figure 4.10 ROC of the proposed RT method and other existing methods
(a) MSRA 1000 (b) MSRA 5000 (c) SED 1 (d) SED 2 (e) SOD
4.5.2.2 Objective evaluation

The objective evaluation of the proposed method is done by ROC curves. The definition of ROC is given in the section 4.5.2. In addition with that, another set of metrics are found to enhance the objective evaluation. The metrics are precision ($P$), recall ($R$) and F-measure ($F_\alpha$).

The percentage of correct assignment of salient pixels is termed as precision $P$. The corrective detection of salient pixel in accordance with ground truth is termed as recall $R$. The harmonic mean between precision and recall is termed as F-measure $F_\alpha$. All these parameters are found for the proposed method and the comparative methods. The binary images are obtained with the help of the otsu automatic threshold algorithm (Gonzalez RC et al. 2004) and mean value of saliency map. Figure 4.10 and Figure 4.11 show the ROC curves and PR charts of the proposed method RT and existing methods for the different datasets. It is undetstood from the Figures 4.10 and 4.11 the proposed method produces higher performance compared to other comparative methods.

The area under ROC curve is shown in the Figure 4.9. The area under curve AUC yields consistently high performance of the proposed method in all the bench mark datasets. MSRA 1000 database scores high AUC for all the extant methods. But the SOD database scores low AUC for all the extant methods. Hence the SOD remains a challenging one for saliency detection.

Mean Absolute Error (MAE) is taken as another objective evaluation test criteria. It gives more balance comparison between continuous saliency map and the binary ground truth (Perazzi et al. 2012). If the salient regions are uniformly highlighted by the saliency map then MAE value is very small. The dissimilarities between the saliency map (prior to thresholding) and
binary ground truth are given by MAE. Figure 4.12 shows MAE of the proposed method and state-of-the-art methods. It highlights the proposed method RT significantly outperforms the other state-of-the-art approaches in terms of MAE measure.

Figure 4.11 Precision, Recall, F measure of the proposed RT method and other existing methods (a) MSRA 1000 (b) MSRA 5000 (c) SED 1 (d) SED 2 (e) SOD
The subjective and objective evaluation are performed for the proposed method. It is compared with ten state-of-the-art methods also. The overall performance of the proposed RT method is good under both subjective and objective analysis and the outcomes are reliable. The accuracy or efficiency of the proposed RT method purely depends on the applications one may choose.

4.5.3 Failure Analysis

The proposed Ripplet transform based visual saliency detection is experimented and the performance of the method is analyzed through various performance metrics with the comparative methods. However, there are some input images which are very challenging to this experiment. Whenever the input image contains the salient objects whose backgrounds are also in the same color, then the saliency is wrongly highlighted. Figure 4.13 shows some of the failure case examples and their ground truth. In these images the backgrounds and salient objects are confusingly blended with one another. Such image cases are not effectively handled by the proposed RT method as
well as state-of-the-art methods. This type of problem will be sort out in the future.

Figure 4.13 Examples of failure cases  (a)Input Image  (b) Ground Truth  (c) saliency map of the proposed RT method

4.6 SUMMARY

To summarize this chapter, an efficient visual saliency detection method using Ripplet transform is proposed. It focuses on detecting the salient regions in the transform domain. The novelty of the proposed method lies on the usage of Ripplet transform in the visual saliency detection. The Ripplet transform can represent the directional singularities very well. The feature maps of various scale and direction are obtained through Ripplet transform. The global and the local saliency maps are obtained from the feature maps by considering the global probability density distribution and feature distribution of local areas. Ripplet transform based visual saliency detection model has better performance compared with the ten state-of-the-art models on five benchmark datasets. As it is a transform domain based visual saliency detection, the mean absolute error is also reduced much. The application of saliency detection models are discussed in the next chapters.